

## Statistical Analysis Selector

**Muhammad Akram Rahman<sup>1</sup>, Aini Jaapar<sup>1</sup>, Wan Zuraida  
Wan Mohd Zain<sup>1\*</sup>, Mazlina Mahdzar<sup>1</sup>, Shahril Azwan Abd  
Rahim<sup>1</sup>**

<sup>1</sup>Pejabat Program Pra Pendidikan Tinggi,  
Universiti Teknologi MARA, Shah Alam, 40450, MALAYSIA

\*Corresponding Author Designation

DOI: <https://doi.org/10.30880/mari.2021.02.03.057>

Received 05 September 2021; Accepted 05 October 2021; Available online 15 December 2021

**Abstract:** Pejabat Program Pra Pendidikan Tinggi (PPT) UiTM is an office that concentrates on programs that enhance students' education at pre-diploma level, which enables them to get entries into Diploma programs. Thus, student's data play an important role for the office to make decisions that can benefit the students. Since statistical analysis has been important in decision making, but no one with a permanent role at the office has a strong Statistics background. It can be quite cumbersome to anyone in the office to find appropriate statistical analysis that can serve office business requirements. A software application was innovated to make effective decision making in suggesting some statistical analysis which can be appropriately done based on given data input. End users will have to input a dataset and then a list of variables can be chosen from for analysis. Next, a few analyses will be given to be chosen from and the end users will get the outcomes that are required. Technically, the app can help to simplify the analysis process by checking the requirements of the statistical analysis. Then, it can give a list of suitable analysis can be picked from and thus results can be obtained in a short time. Agile methodology concept was carried out which briefly involved of technical analysis plan and design phase for fulfilling conditional requirements between statistical analysis and application utilities, application development phase for building the application and finally evaluation phase for the app to be assessed by PPT officers. With this innovation, there will be less headache and time consumption on exploring and doing some research about many statistical methods that will not even be used. Thus, at the very least, PPT officers of non-Statistics background have no need to do any research on statistical analysis, but necessary outcomes of business objectives can be attained from the software application.

**Keywords:** Corporate Social Responsibility, Statistical Analysis, Automation, Visual Studio

## 1. Introduction

Pra Pendidikan Tinggi (PPT) program is for Sijil Pelajaran Malaysia (SPM) graduates who acquired SPM results that do not comply with the requirements specified by UPU. PPT program is open to applicants from both B40 and non B40 families to further in higher education studies. To date, more than 43,000 students has been given educational opportunities from 2010 until 2020. PPT program helps to improve academic performance through pre-commerce and pre-science programs which are based on UiTM pure ethical values and professionalism.

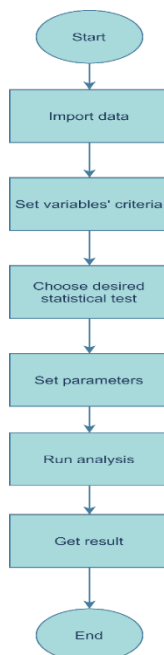
With the visions and missions in hand, accurate and efficient decision making should be an aspect that is really important to the organization. This can bring impacts to many resources especially in financial aspects. Less resources spent for a better goals achieved should be a good rule of thumb [1].

## 2. Problem Statement

Being a critical matter to achieve the goals, statistical analysis has been important in decision making, but no one with a permanent role at Pra-Pengajian Tinggi (PPT) Office has strong Statistics background. It can be quite cumbersome to anyone in the office to find appropriate statistical analysis that can serve office business requirements. Furthermore, a lot of time is needed just to go through the learning process regarding statistical analysis. Not to mention the time that would be wasted from excluding many useless analysis that is not suitable to business goals in hand [2]. In addition, MS Excel is the only familiar tool to them to at least manually keying in formulas for data analysis, given that statistical tools in MS Excel requires strong foundation in Statistics and this may even take more time to execute the analysis.

Yet many kinds of technologies have been developed to work with analysis for business purposes in fast pace, this however does not solve the problem quickly as the foundation in statistics to use the technological tools should be strong in the first place [3].

Nowadays, there are many statistical tools available for researchers to use. For most of the tools, the common ways of executing the analysis are shown as in Figure 1 below.



**Figure 1: Common steps needed to run analysis using most statistical tools**

The main issue with the conventional method used as in Figure 1 is at the step of choosing desired statistical test. The users need strong foundations in statistics to be able to decide on the appropriate tests to perform. Basic knowledge on just a few tests would not be enough as many assumptions and measures need to be taken into account to use the most suitable test because dataset can be of various kinds and so do the statistical tests. This would lead to a more time consumption needed to learn statistical analysis.

### 3. Objectives

The objective of this project was to build a software application to deal with the mentioned problem of statistical tools' users having weak background in Statistics. There were a few goals that were intended to be achieved from the development of the app. First, the app was intended to ease up PPT operations in executing tasks regarding data analysis. The analysis processes regarding PPT dataset are usually sophisticated as to why they need to be facilitated by this app. Next, as more time needed to handle the manual and slow analysis processes as for the users, maybe just by using MS Excel program, this app was made to quickly generate results of statistical analysis regarding PPT dataset in seconds [4]. Lastly, the software app was made to be installed and used in several kinds of machines especially on laptop to facilitate working with mobility. In addition, it must also have offline feature so the requirement of internet access subsequent to its installation would not be a problem.

### 4. Methodology

As this is mainly involve software development, agile methodology (Figure 2) was used to cater with the evolution of demands and solutions in the development process. Many things would be added from time to time throughout the process.



**Figure 2: The Agile methodology steps carried in this project**

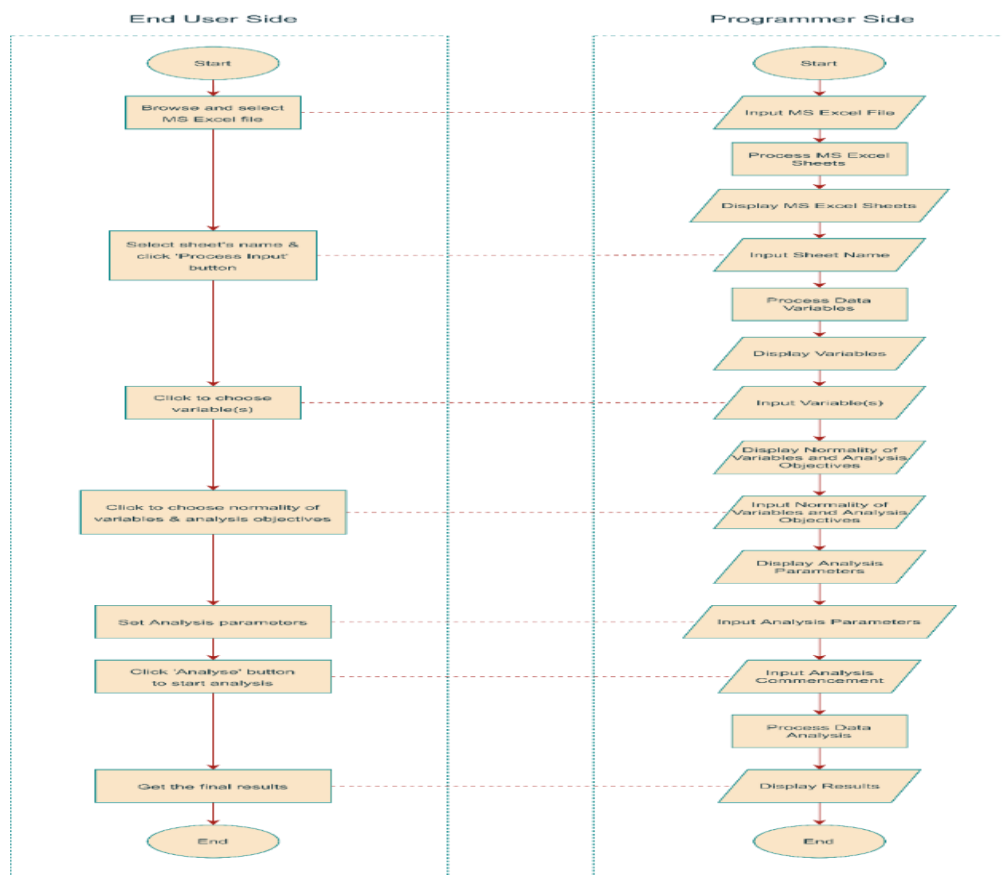
Using this methodology, business requirements were gathered first before proceeding to the planning phase. With the requirements obtained, they were planned for their suitable arrangements. A requirement may comes after another or vise versa, and this needs really plenty of time to make it looks logical and making sense.

Next, prior preparation materials were then matched with design functionality. In the software utility library there are many things that need to be cater of carefully because some software utilities like textboxes, buttons, panels and many more may not work correctly according to our plan.

Development phase comes after all materials and designs have been taught of to work together smoothly. As this is a statistical analysis app project, it requires 3<sup>rd</sup> party library for statistical calculation purposes and this sometimes may not work well with the design plans that have been intended to work together in this development phase. This development phase was the trickiest part because many errors would be found here and debugging need to be done several times as to solve the unmatched design and business requirement. Later, the developed software app was deployed for real usage by the end user. Although this indicates that the app would work well, however, many errors could still be prone to happen for some cases and even more time must be spent just to find out solutions to a few errors of unknown causes.

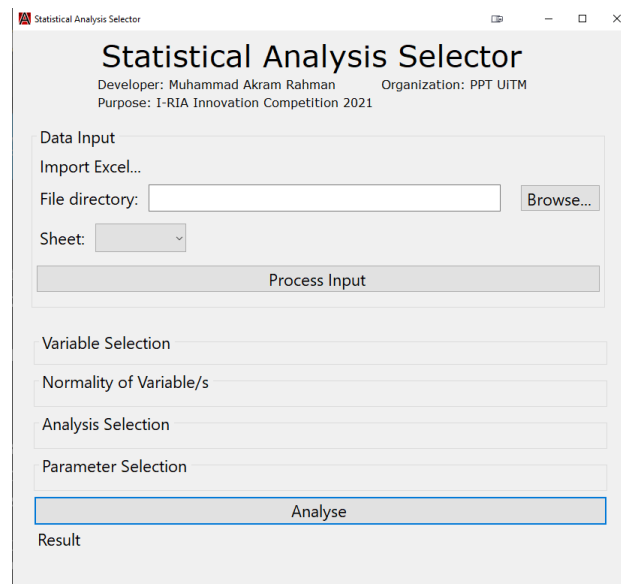
Furthermore, prior to deployment phase, the app still needs to undergo evaluation phase where more errors would be checked in broader aspects where the app may need any adjustments or alterations to work better in end users' machine in the future. For the development of the software application in this project, extraction of the requirement of sophisticated knowledge to conduct statistical analysis has been a primary focus. The end users would not have to deal with the learning of various statistical analysis prior to using this new application.

Figure 3 illustrates the ways the application would work from two different perspective, one from the end user and another one from programmer's view. The most valuable concept of this application is at the analysis selection step where conventional statistical tools are difficult to use for any layman. After providing necessary information on the required variables and their normality, the new software application will compute in the background for suitable analysis that can be used by the users. In addition, the analysis selection would not be shown as in the names of any statistical test but rather the objectives of the analysis. If the end users choose to bring in two variables, the analysis selection will be the objectives of finding relationship between both variables or their differences according to their dependencies to each other for being related or independent.



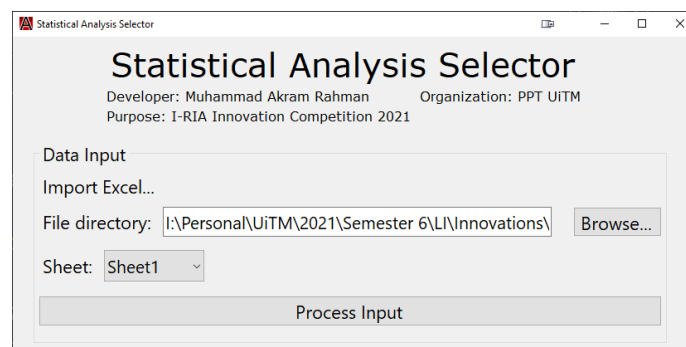
**Figure 3: Proposed pseudocode for the new software application**

## 5. Findings and Discussions



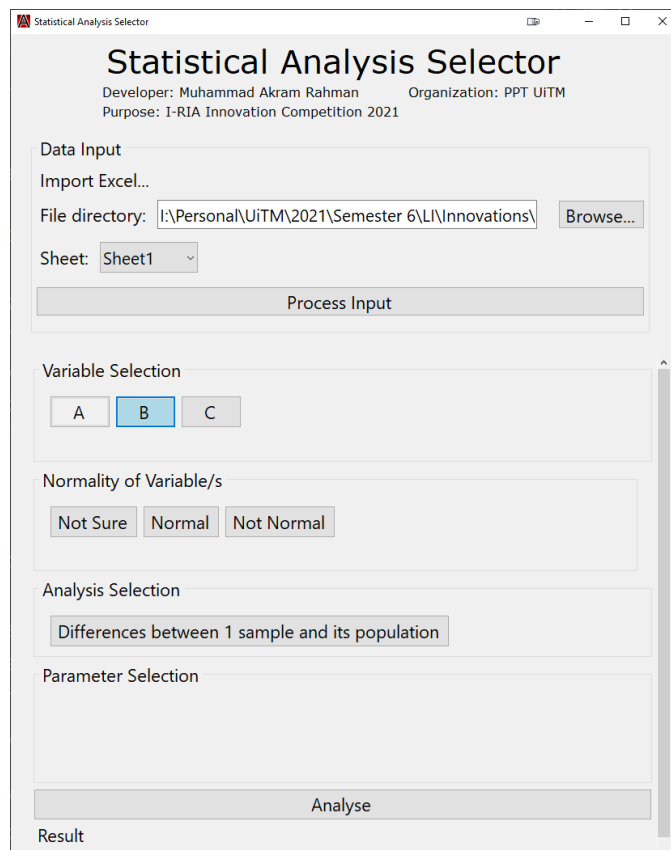
**Figure 4: The built software application**

This project has come out with the development of the software application as intended. Briefly, this end user for this application will firstly have to locate the directory of the data source file as in MS Excel .xlsx format after clicking the ‘Browse...’ button (Figure 4). After that, the user has to choose the Excel sheet that contain the required data. Next, the button namely ‘Process input’ can be clicked to start the analysis.



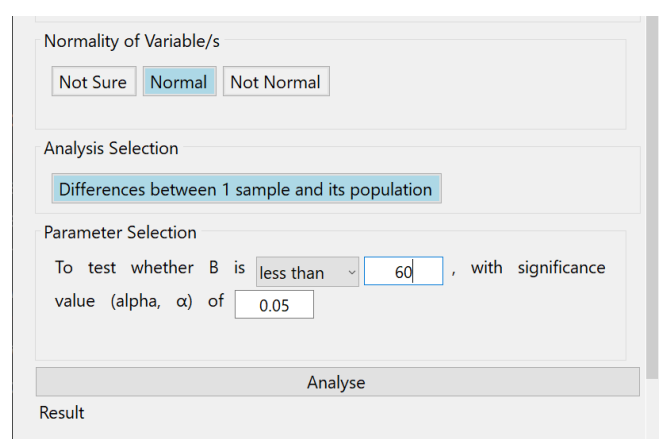
**Figure 5: The sheet's name can be chosen**

After the last button click, selections of variables from the chosen dataset will be shown to be chosen from (Figure 5). In case of many variables exist in the dataset, the end user can select any number of variables of interest. As a matter of fact, the number of chosen variables will affect further processes in the background of the application where different statistical analysis will be displayed as options. Before the analysis of choice, options indicating the normality of the data will be shown too as to let the end user to let the program knows about the normality of the variables. There are 3 choices, normal, not normal and unknown. The first two choices will objectively let the program show further options of analysis that are related only to either of these first two choices as statistical tests can be used on either normal or non-normal data; and for the last one, the program will compute the normality of the variables of interest, in which the program will decide the normality of the variables. This can give a freedom to the end user to decide on the normality of the variables or just let the program decides.



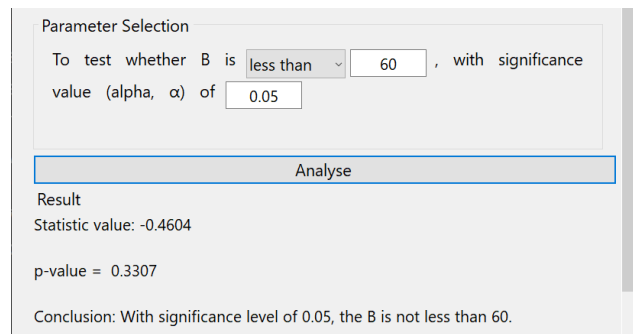
**Figure 6: Variables selection, normality and analysis selection**

Later, under the ‘Analysis Selection’ part, a few options of statistical analysis will be shown (Figure 6). In this matter, the options of analysis are not shown as in any name of statistical test but rather the objectives that can be achieved from the variables are displayed. By doing this, any sophisticated or unfamiliar names can be abstracted from the end user and just let the program makes decisions on the appropriate tests.



**Figure 7: Parameter setting**

After the analysis selection, a few settings on the parameter of the analysis need to be set (Figure 7). The parameters can be the alpha value, tested value for a single sample analysis and many more. This part also gives freedom to the end user to alter the parameter values as to get desired result after clicking the ‘Analyse’ button.



**Figure 8: Result of analysis in the software application**

Using the software application, the end user does not need to think anything about statistical analysis tests which could consume more time and resource to learn (Figure 8). Instead, the end user just has to know the objectives that need to be achieved from the imported dataset. This mainly benefits from the abstraction on the statistical analysis done in the background of the program.

With this innovation, end users just need less than a minute to generate results from statistical analysis as there are just a few steps that need to be done to get the final results. The steps are easy to follow and no complicated mouse movements and statistical terms to encounter. With regular use of the app, they can even spend less than 30 seconds to get the desired results.

The software application has been distributed to the officers and Most of them were happy with it, in which they verbally rated for more than 80% satisfaction level. This shows that the app can be justified to serve their business purposes by encountering previous problem regarding time-consuming data analysis.

Afterall, the software app made can be easily used by anyone in the office without the requirement of strong Statistics background as the multi-platform app can abstract or put aside complex analysis processes which are done in the background.

## 6. Conclusion

The development of the software app has met all the objectives of this project. It really serves well its users of non-Statistics background where sophisticated statistical analysis processes are be handled smoothly in the background. Without encountering those difficult processes by the users, this software app can be considered as a non-Statistics friendly app, which it can quickly generate desired analysis results with no hassle.

Meanwhile, this app was built to work well and can be installed in most computers especially laptops. With every feature built into the package, it can also be used offline where this will not hinder anyone from using it anywhere especially without the difficulty of internet network problem.

It is recommended that this app can be further developed as a mobile application where it gives more flexibility as it can also be used with mobile phones to generate analysis results. Furthermore, more statistical library for software development should be explored or built for a better ones as to work smoother with the statistical app.

## Acknowledgements

Special thanks to all staff and management team Pra Pengajian Tinggi UiTM Shah Alam for continuous supports on working with this innovation project. Finally, special credit to Bahagian Transformasi Universiti (BTU), UiTM for providing valuable dataset for software application testing.

## References

- [1] J. A. S. Salleh, M. Mahdzar and A. Rahim, "The blueprint of pre-higher education program: a platform to uplift the B40 economic position through Islamic social finance," *Contemporary Issues in Islamic Social Finance*, pp. 246-259, 2021.
- [2] D. Brooks, "How Long Should It Require You to Learn Something?," 2017. [Online]. Available: <https://medium.com/swlh/how-long-should-it-require-you-to-learn-something-b13bb6cacd1f>.
- [3] UNCTAD, "Technology and Innovation Report 2021," United Nations Conference on Trade and Development, 2021.
- [4] M. J. T. Byrne, "Comparison of Manual Versus Automated Data Collection Method for an Evidence-Based Nursing Practice Study," *Applied Clinical Informatics*, vol. 4, no. 1, pp. 61-74, 2013.